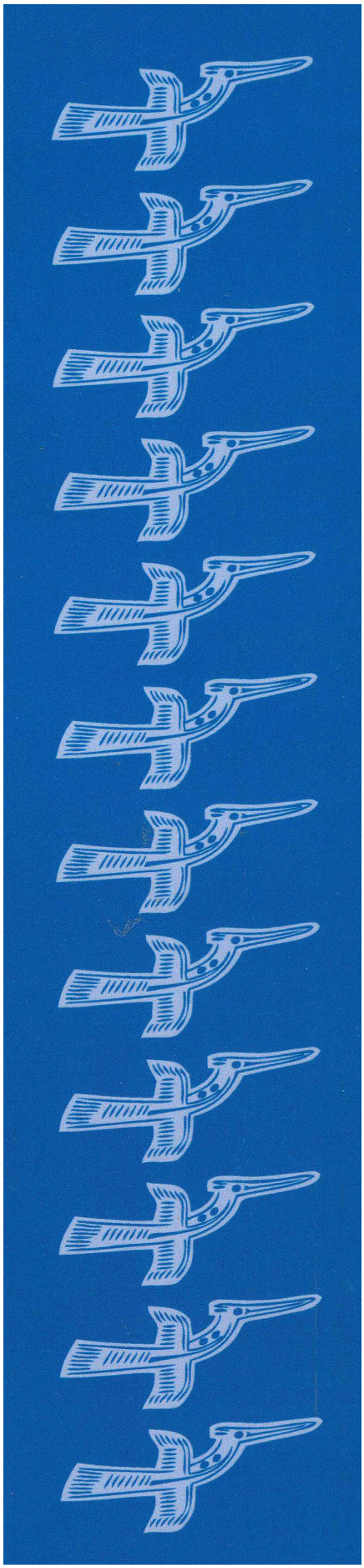
Mon-Khmer Studies

Volume 42



Authors: K. S. NAGARAJA, Paul SIDWELL, Simon GREENHILL
Title: *A Lexicostatistical Study of the Khasian Languages: Khasi, Pnar, Lyngngam, and War*
Pages: 1-11
Date Received: 26/1/2013
Revised Text Accepted: 8/5/2013
Copyright for this paper vested in the authors
Released under Creative Commons Attribution License

Volume 42 Editors:
Paul Sidwell
Brian Migliazza

ISSN: 0147-5207
Website: http://mksjournal.org

Published by:

Mahidol University (Thailand)

SIL International (USA)

A Lexicostatistical Study of the Khasian Languages: Khasi, Pnar, Lyngngam, and War.[1]

K. S. Nagaraja
Retired, Mysore, India

Paul Sidwell
Australian National University

Simon Greenhill
Australian National University

Abstract
This paper presents the results of lexicostatistical, glottochronological, and Bayesian phylogenetic analyses of a 200 word data set for Standard Khasi, Lyngngam, Pnar and War. Very few works have appeared on the subject of the internal classification of the Khasian branch of Austroasiatic, leaving the existing reference literature disappointingly incomplete. The present analysis supports both the strong identity of Khasian as a unitary branch, with an internally nested branching structure that fits neatly with known historical, geographical and linguistic facts. Additionally, lexically based dating methods suggest that the internal diversification of Khasian began roughly between 1500 and 2000 years ago.
Keywords: Lexicostatistics, Bayesian phylogenetics, language classification
ISO 639-3 language codes: kha, lyg, pvb, aml

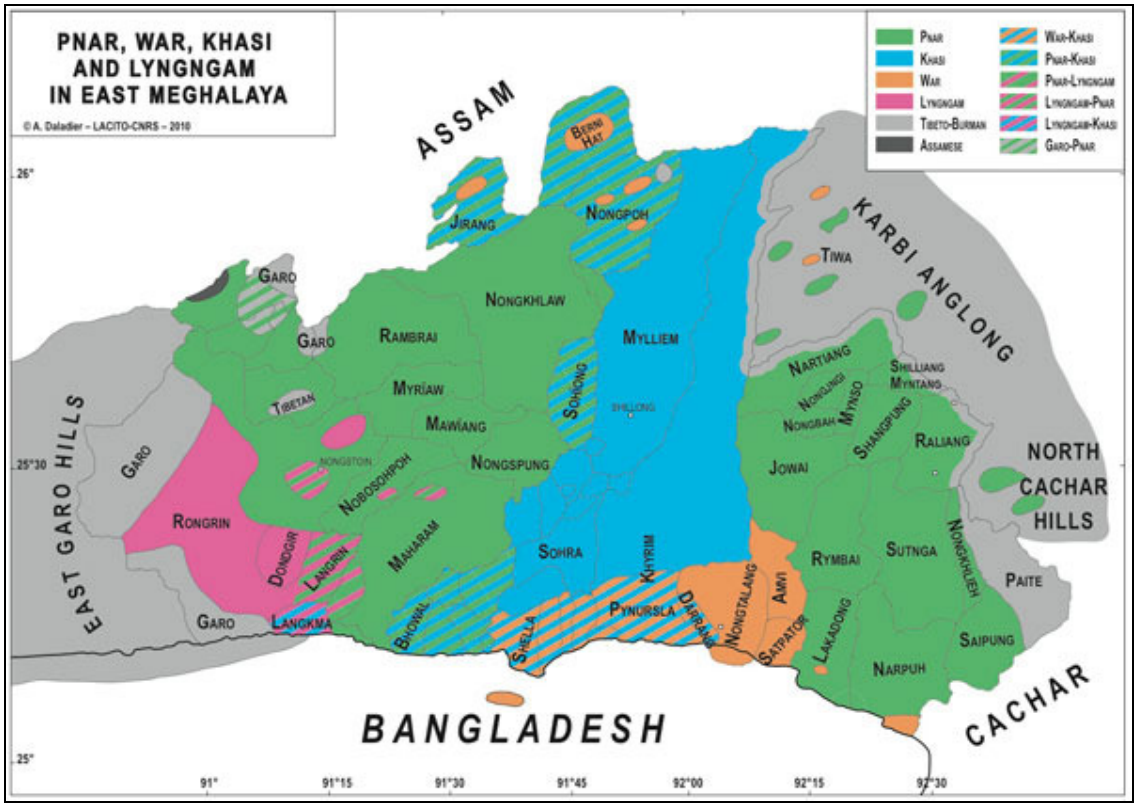


Figure 1: Map of Khasian varieties from Daladier (2010)

[1] The present paper extends a 2004 study by K. S. Nagaraja "A Lexico-statistic study of Khynrian and Lyngngam dialects of the Khasi language" (The NEHU Journal 2.1:43-56). Special thanks are due to Hiram Ring and Mankular Gashnga for assistance with data and analyses in the preparation of this paper.

K. S. NAGARAJA, P. SIDWELL, S. GREENHILL. 2013.
A Lexicostatistical Study of the Khasian Language. *Mon-Khmer Studies*. 42:1-11

1. Background: previous studies

Comparative-historical analysis of Khasian remained underdeveloped through the 20th century, primarily because attention has traditionally focused on the standardized variety, which enjoys official status and widespread use in religious contexts in Meghalaya state. We see an indication that language attitudes were well entrenched already in the 1800s in this extract from Roberts (1891) Khasi grammar:

> In this work, the dialect of Cherrapoonjee is taken as the standard, because it is the purest, as universally acknowledged by the natives, besides being more amenable to systematical arrangement than the patois of the smaller villages.
> (Roberts 1891, xiv)

Robert's text also includes an extensive list of "ugly" (presumably quite popular) non-standard pronunciations that speakers were urged to avoid at all costs. In striking contrast, *The Linguistic Survey of India* (Grierson (1903) correctly recognized four languages which correspond to the four speech varieties analysed in this paper (Khasi (Khynrium), Pnar (Synteng), War, and Lyngngam[2]) and provided some useful comparative lexical and syntactic examples. However, Grierson's data suffered from limitations in the transcription and other gaps that made it difficult to provide a basis for linguistic analyses, and it would be approximately a century before improved data, similarly systematically organized, would start to become available.

Of the present authors, Nagaraja collected data for both Standard Khasi and Lyngngam in 1988, and published a paper on the status of Lyngngam in 1996. That paper made various observations on the grammatical, lexical and phonological correspondences between Khasi and Lyngngam, including the important observation that, "around forty percent of Lyngngam's vocabulary seems to be unrelated to Khasi." The same author followed up with a lexicostatistical study in 2004, based on a 200 word list that was subsequently used as the basis for our more recent and extensive analysis that is the main topic of this paper. That study found 43.9% of cognates between Khasi and Lyngngam, and using Lee's (1953) glottochronological method calculated a separation date of 1,890 years. Subsequently, other lexicostatistical studies have been conducted independently.

Brightbill et al. (2007) conducted a sociolinguistic survey of Khasian villages in Bangladesh, and at a couple of locations within Meghalaya, presenting their wordlists and lexical analyses in their online report (see URL in the references). Although focusing on the War varieties within Bangladesh, Brightbill et al. provided useful lexical data for a number of Khasian varieties, in the form of a comparative lexicon with more than 300 items. On the basis of that list they calculated the lexicostatistical matrix reproduced as Fig. 2 ("lexical similarity chart" in their terminology):

Magurchora (I)										
92	Barenga (E)									
90	91	Niralapunji (A)								
90	89	88	Singur (D)							
90	91	88	88	Aliachora (B)						
88	86	88	86	87	Dabolchora (C)					
89	86	85	87	87	83	Amlarem (J)				
35	35	35	35	35	34	36	Noksia (F)			
18	18	18	18	18	17	17	31	Lyngngam (H)		
33	33	32	34	32	31	34	49	30	Jaintiapur (G)	
25	25	26	26	25	24	28	51	34	48	Shella (K)
29	29	29	30	29	27	32	55	36	53	75 Shillong (L)

Figure 2: Lexicostatistical matrix of selected Khasian varieties from Brightbill et al. (2007:17)

[2] Another apparent substantial Khasian speech community is Maram, to the west of the main Khasi area, corresponding to the green Pnar (!) area on the western side of Daladier's map. Maram is not treated here due to lack of suitable data, but we can report that impressionistically it is very similar to Standard Khasi.

In Brightbill et al.'s scheme the wordlists are identified mostly by place names: the first six above are War varieties spoken in Bangladesh, while *Amlarem* is a War dialect from Meghalaya, the *Noksia* and *Jaintiapur* are Pnar varieties, the *Shella* is ambiguously explained as being "Khasi-War", and the *Shillong* is from a speaker of Standard Khasi. The main result is that War lects in Bangladesh are clearly identified as varieties of one language with percentages all above 80%. However, the other figures are more difficult to interpret, especially in respect of the particularly low percentages that Lyngngam shares with other lists - as low as 17% - well below what we might anticipate given the analysis of Nagaraja (1996). Their calculations appear to be heavily skewed by a failure to allow for missing items in the lists compared, and are included here mainly for the sake of completeness in reviewing the lexicostatistical data on Khasian.

Another of the present authors, Sidwell, attempted his own lexicostatistical study of Khasian, which is presented in his (2009) survey of Austroasiatic classification. That study used the standard 100 word Swadesh list, aggregating items from the following sources:

- Lyngngam data from Nagaraja (1996),
- Khasi from standard dictionaries,
- Amwi from Weidert (1975),
- Pnar (Noksia) and War (Amlarem) from Brighthill et al. (2007).

Cognates were identified manually and a matrix generated (Fig. 3) automatically using Jacques Guy's Glotpc.exe program:[3]

Lyngngam				
63	Khasi (Shillong)			
54	75	Pnar (Noksia)		
41	55	57	War (Amlarem)	
37	53	51	80	Amwi (Weidert)

Figure 3: Lexicostatistical matrix for five Khasian varieties, by Sidwell (2009)

The above figures were interpreted as indicating that the languages fell into two sub-groups: War versus a Khasi-Pnar-Lyngngam group, with the latter having an ambiguous structure. Generally the main finding that the War varieties form a distinct sub-branch is supported strongly by comparative phonology. War is strongly marked by historical vowel restructuring that saw many mergers with high front vowels, and dissimilatory restructuring of diphthongs. Some examples can be seen in the following table (Fig. 4) of data extracted from Lyngngam from Nagaraja (2004), Brightbill et al. (2007), and Amwi from Weidert (1975).

Gloss	**Lyngngam**	**Khasi (Shillong)**	**Pnar (Noksia)**	**War (Amlarem)**	**Amwi**
'two'	*aːr*	*ʔaːr*	*ʔaːr*	*ʔɨ*	*ʔũ*
'chicken'	--	*ʔiar*	*ʔiar*	*sɨʔi*	*sʔi*
'fish'	*kʰa*	*kʰa*	*kʰa*	*hi*	*hi*
'red'	*ənsaw*	*saw*	*sao*	*sɨa*	*sia*
'stone'	*maw*	*maw*	*mao*	*ʃmɨa*	*ʃmia*

Figure 4: Comparative data illustrating phonological innovations in War

It is apparent that the lexicostatistical studies conducted so far have been very limited in scope, and conducted with differing data sets that make their result difficult to compare and assess. In this context it was decided to extend Nagaraja's (2004) study, by adding data representing Pnar

[3] Figures on branches are words retained per 1000.

and War to the 200 word list already used for Lyngngam and standard Khasi, and additionally to add Palaung data - Palaung representing a more distantly related Austroasiatic language - to securely root the tree and test overall coherence of Khasian.

2. The present study

The present study takes the data set of Nagaraja (2004) to which are added data items for:

- Pnar, Jaintia dialect from Ring (2012).
- War, Lamin dialect from Gashnga (forthcoming).
- Palaung, Namshan dialect from Shorto 2013.

All the data are provided in a table as an appendix to this paper. Cognates are scored in the rightmost column of the table using letter codes, according to the method specified by Guy (1994) in which members of the same etymon are given the same letter, loans and empty fields are given *. Nagaraja's (2004) cognate assignments were reassessed in the light of the new data, and Sidwell's ongoing proto-Khasian phonological reconstruction,[4] resulting in some changes. The scores were then processed with Guy's GLOTPC.EXE to count the pair-wise percentages, generating the table at (Fig. 5):

	Khasi	Pnar	Lyngngam	War	Palaung
Khasi		74	62	52	20
Pnar	74		55	54	19
Lyngngam	62	55		41	18
War	52	54	41		20
Palaung	20	19	18	20	

Figure 5: Lexicostatistical table for Khasi, Pnar, Lyngngam, War and Palaung.

Overall the matrix indicates straightforward nesting branching relations within Khasian, and unambiguous rooting based on the strikingly consistent 18~20% cognacy with Palaung. Further analysis with Guy's GLOTED.EXE indicates that the real percentages diverge from theoretically predicted percentages by no more than 2% in respect of any pair-wise comparisons, so we can have some confidence that the analysis is not significantly distorted by drastic differences in rates of lexical change. As regards to inferring interference by borrowing, it appears that the cognacy rates with War are indicative; we know that War and Lyngngam speakers are geographically separated, and can assume that the figure of 41% counted between them is not significantly distorted by loans. On the other hand, Khasi and Pnar show higher percentatages against War (52% and 54% respectively) and the somewhat higher agreement between Pnar and War, which are known to be in contact, is surely indicative of some mutual borrowing, which has not been identified and scored so in our dataset. Thus, although the pair-wise comparions of Pnar-Lyngngam and Pnar-War show similar values (55% and 54% respectively) we can assume that the latter figure is likely to be high because of undiagnosed borrowing (as borrowing between Pnar and War is far more likely than between either and Lyngngam[5]). Similarly, the higher agreement between Khasi-Lyngngam (62%) versus Pnar-Lyngngam (55%) is likely to be indicative of some borrowing of Standard Khasi words into Lyngngam. Of course, it must be acknowledged that it is possible that the above patterns are largely the result of differences in rates of change, but logically it is difficult to see how that would produce such a tidily branching nested tree, as opposed to a more random pattern.

The figure of 74% agreement between Khasi and Pnar is strikingly consistent with 75% figures independently obtained by Brightbill et al. (2007) and Sidwell (2009), and provide significant comfort to the view that they are more or less indicative of the real distance between the two languages. The figures indicate that a high degree of mutual lexical intelligibility is to be predicted, approaching the threshold for treating them as dialects of the same language. Clearly

[4] At the time of writing a 2012 version of this reconstruction is available online at sealang.net/monkhmer. It is expected that this will be replaced with an extensively revised version later in 2013.

[5] If anything, all three are likely to share unrecognised loans from Standard Khasi.

Khasi and Pnar sub-group closely, Lyngngam then appears to sit above Khasi-Pnar, and all three are more distantly related to War, which (as noted above) also is known to have a divergent phonological history.

For those who are bold enough to pursue the question, it is also possible to apply glottochronological calculation to our figures, in full awareness of the harsh critiques of glottochronology, especially since Bergsland and Vogt. (1962). We have done so, using Lee's (1953) formula **t = log C / (2 log r)** and his retention rate for the 200 item list of 80.5% per thousand years. Applying this formula to the lowest pair-wise percentage at each apparent node, we get the following tree with divergences dated in years (y) before present at Fig. 6.

```
Khasi ------:74%/694y----:55%/1378y----:41%/2054y----:18%/3951y
Pnar -------'             |             |             |
Lyngngam ----------------'              |             |
War ------------------------------------'             |
Palaung ----------------------------------------------'
```

Figure 6: Family tree with glottochronological dating of divergences table for Khasi, Pnar, Lyngngam, War and Palaung.

Further computational analyses were carried out on the dataset. Firstly a neighbor net was generated using SplitsTree v4.11.3 (Bryant & Moulton 2003) by Simon Greenhill, here at Fig. 7. The result, displayed below, is quite straightforward, and is largely consistent with the lexicostatistics: the close relation between Khasi and Pnar is reproduced, and the marginally closer relation of them to Lyngngam versus War is evident. No disproportionate inferring signals are evident.

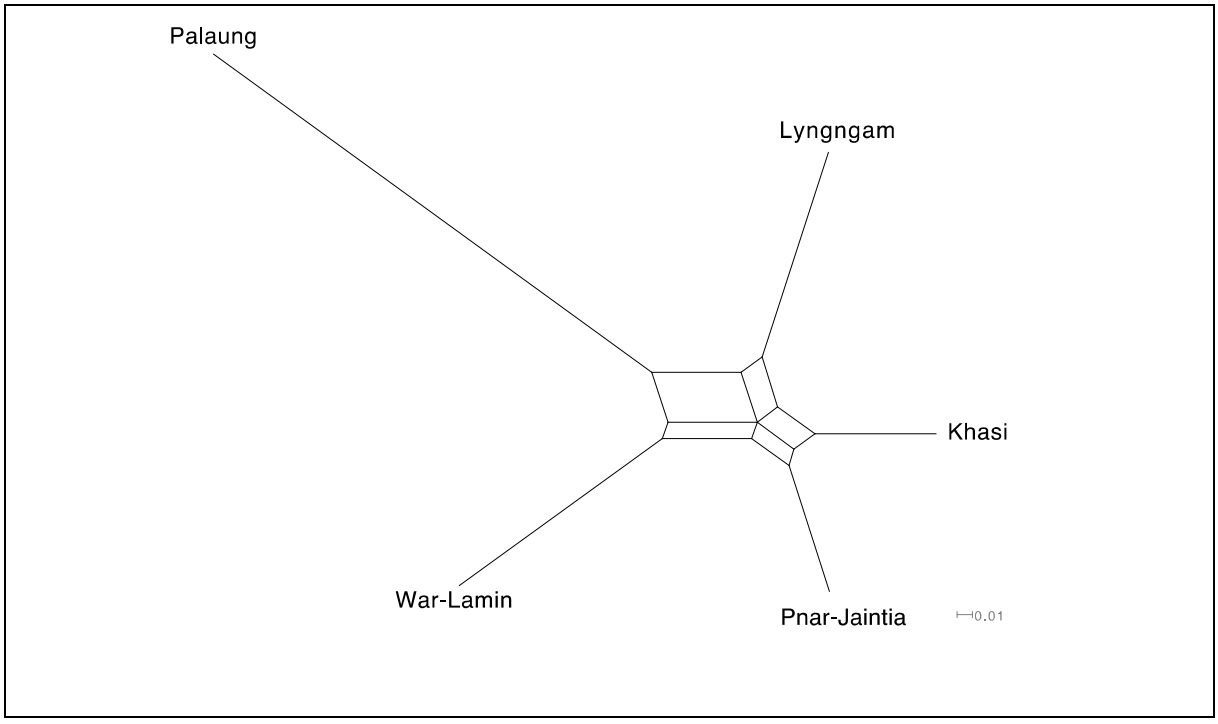


Figure 7: Neighbor Net for Khasi, Pnar, Lyngngam, War and Palaung.

Next, a Bayesian Phylogenetic analysis was run by Greenhill using BEAST v1.7.4 (Drummond et al. 2012). Here, a simple Continuous-Time Markov Chain model was used to analyse the binary presence/absence of cognates implementing a strict clock for inferring rates of cognate gains and losses. The analysis was run for 2,000,000 generations, sampling 1,000. The first 200,000 generations were discarded as "burn-in" after inspection of the traces showed that this was sufficient time for the chain to stabilize (c.f. Greenhill, in press). The results are similarly consistent with the lexicostatistics; the number 1 at each node indicates 100% probability of the

branching, as the program consistently generated the same tree with every pass through the data. Additionally the tree is constrained to indicate a time depth of 500 years BP for the Khasi-Pnar split, for the sake of generating a calibrated tree. In so far as we are able to offer any objective bases for calibrating the chronology, the Buranji chronicles of the Ahom kingdom apparently reference the Pnar kingdom at Jaintia about 500 years before present, which suggests a floor under the separation of Khasi and Pnar (e.g. Gait 1906:255 lists Jaintia kings from approximately 1500 AD onwards). In this context, the glottochronological calculation of 694 years for Khasi-Pnar separation seems quite realistic, although still admittedly speculative. The Bayesian analysis estimates the age of the Khasi-Pnar split to be similar – but younger – with a mean of 535 years (95% Highest Posterior Density Interval = 500-603 years), see Fig. 8. In terms of the age of the Khasian subgroup itself, glottochronology estimates the age at 2054 years, while the Bayesian analysis places the origin of this subgroup at a younger median of 1350 years (95% HPD = 1028-1737 years). Given that the Bayesian dating estimate is calibrated to the youngest possible age indicated by our meager historical sources, the estimate of 1350 years is quite likely to be an underestimate, although as such it establishes a reasonable minimum parameter for speculations about pre-Khasian migration into Northeast India.

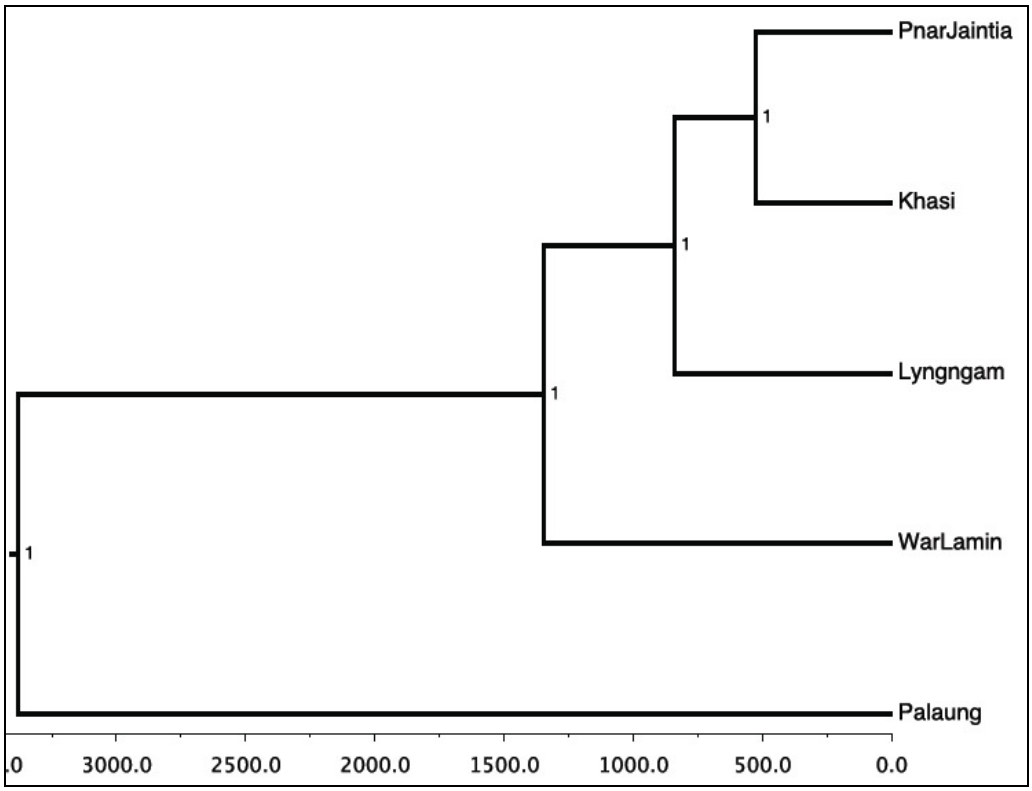


Figure 8: Bayesian Phylogenetic analysis for Khasi, Pnar, Lyngngam, War and Palaung.

3. Concluding remarks

The present study makes a further contribution to the emerging field of comparative Khasian linguistics, with a quantitative analysis of lexical correspondences that supports both the unity of the Khasian branch, and a strong nested internal structure. Within Khasian, the War language(s) form the highest branching node, consistent with indications of historical phonological restructuring. The remaining languages form a tightly linked subgroup, with Lyngngam placed outside a Khasi-Pnar core. Whilst these results are intriguing, fine-grained lexical, grammatical and phonological analyses should be applied to further rigorously infer the subgrouping of the Khasian languages.

References

Bergsland, Knut, Hans Vogt. 1962. On the Validity of Glottochronology. *Current Anthropology* 3.2:115-153.

Brightbill, Jeremy, Amy Kim, and Seung Kim. 2007. *The War-Jaintia in Bangladesh: a sociolinguistic survey*. SIL Electronic Survey Reports.

Bryant, David & Moulton, Vincent. 2003. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2), 255–265.

Daladier, Anne 2010. (web resource) Carte du groupe pnarique-war-lyngngam http://lacito.vjf.cnrs.fr/themes/tibeto-birman/images/East-Meghalaya-PWKL600.jpg

Drummond, Alexei J., Suchard, Mark A., Dong, Xie, & Rambaut, Andrew. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8), 1969-1973.

Gait, E. A. 1906. *A History of Assam*. Calcutta: Thacker, Spink & Co.

Gashnga, Mankular. forthcoming. *Lamin(War)-English Dictionary*. Canberra, Asia-Pacific Linguistics.

Greenhill, Simon J. In Press. Evolution and Language: Phylogenetic Analyses. In *The International Encyclopedia of the Social and Behavioral Sciences*. Wright, JD (Ed). Elsevier.

Grierson, G.A., 1903, *Linguistic Survey of India*. Vol. (iii) Part (ii) (Reprint 1967). Delhi. Banarashidaas.

Guy, Jacques. 1994. GLOTTO (web resource) http://linguistlist.org/issues/5/5-630.html

Lees, Robert B. 1953. On the basis of glottochronology. *Language* 29: 113-127.

Nagaraja, K.S. 1996. Status of Lyngngam, *Mon-Khmer Studies*, Vol. 26: 37-50.

Nagaraja, K.S. 2004. A Lexicostatistic study of Khynriam (St.Khasi) and Lyngngam dialects of the Khasi language. *The NEHU Journal* II.1:43-56.

Ring, Hiram. 2012. A phonetic description and phonemic alanlysis of Pnar. *Mon-Khmer Studies* 40:133-175.

Roberts, H. 1891. *A Grammar of Khassi language*, London.

Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich, Lincom Europa.

Shorto, H. L. 2013. *Palaung Word List: Based on materials collected from Pan Shwe Kya, Namhsan, Sept-Oct, 1957*. Canberra, Asia-Pacific Linguistics.

Weidert, Alfons K. 1975. *I tkong Amwi. Deskriptive Analyse eines Wardialekts des Khasi*. Wiesbaden: Harrassowitz

Appendix: Lexical data and cognate scores

Lexical data and cognate scores (* marks loanwords and missing forms)

	Gloss	Khasi	Lyngngam	Pnar/Jaintia	War/Lamin	Palaung	Cognate scores
1	all	*roʔ*	*prok*	*warɔʔ*	*bərɒʔ*	*paj*	aaaab
2	and	*baːd*	*nam*	*wa*	*waː*	-	abcd*
3	animal	*mraːd*	*mraːd*	*mraːd*	*mrat*	*to*	*****
4	ashes	*dpej*	*əpaw*	*tpai*	*tvɔ*	*kəhvaŋ*	abacd
5	at	*ha*	*he*	*ha*	*ti*	--	aaab*
6	back (anat.)	*dien*	*bad don*	*rɲkʰi*	*təmpɒŋ*	*krɔŋ*	aabcd
7	bad	*sɲiew*	*kɨncʰa*	*siʔ*	*kɔ̃m*	*kʰu (?)*	abcde
8	bark (of tree)	*snep*	*snieʔ*	*sneiʔ*	*snɪəʔ*	*gɔʔ*	aaaab
9	because	*namar*	*amte*	*neibʰaʔ*	*kaʔ*	*(?)*	abcd*
10	belly	*kpoʔ*	*ləwbaʔ*	*kpɔʔ*	*pɒʔ*	*vɛʔ*	abaac
11	big	*heʔ*	*kɨmba*	*hɛʔ*	*mɪa*	*daŋ*	ababc
12	bird	*sim*	*sim*	*sim*	*ksem*	*sim*	aaaaa
13	to bite	*daʲt*	*kɨnnap*	*dait*	*hit*	*gaʔ*	abacd
14	black	*joŋ*	*ɨŋŋoŋ*	*jɔŋ*	*priŋ*	*jəm*	aaabc
15	blood	*snam*	*snam*	*snam*	*rnʊə*	*hnam*	aaaaa
16	blow	*pɨrsat*	*pʰinnur*	*slu*	*pet*	*put*	abcdd

17	bone	*ʃʔeŋ*	*cʔeŋ*	*ʧʔeiɲ*	*ʧiəʔ*	*kəʔaŋ*	aaaaa
18	breathe	*riŋ mɨnsiem*	*riŋ ɨnsom*	*rɪŋ mɲseim*	*reŋ hənsʊ*	*pʰum*	aaaaa
19	burn	*tʰaŋ*	*ɨntʰaŋ / tʰɨnnəŋ*	*tʰaŋ*	*ɵaŋ*	*gut*	aaaab
20	child	*kʰun*	*kʰon*	*kʰɔn*	*hʊn*	*kuən*	aaaaa
21	cloud	*lʔoʔ*	*lʔoʔ*	*lʔɔʔ*	*ləmpem*	*ut*	aaabc
22	cold	*kʰriat*	*bɨnsir*	*kdʒam*	*ktjam*	*kat*	abccd
23	come	*wan*	*lɨnnar*	*wan*	*va, van*	*hluh, rɔt*	abaac
24	count	*ɲjaw*	*cʰɨnnan*	*niaw*	*ʧaː*	*dir*	abacd
25	cut	*kʰap*	*kʰɨnnap*	*aʔ*	*pam*	*set*	aabcd
26	day	*sŋi*	*sŋej*	*sŋi*	*ʃŋaː*	*səŋi*	aaaaa
27	die	*jap*	*ɲinnap*	*jap*	*jip*	*jəm*	aaaab
28	dig	*tiʔ*	*tɨnniet*	*tiʔ*	*tɪəʔ*	*puər*	aaaab
29	dirty	*ɟaboʔ / ɟakʰlia*	*dɨmmut / ɟɨmbaʲt*	*taroj*	*ʧəmet*	*ɲu ɲu (?)*	abcde
30	dog	*ksew*	*ksu*	*ksaw*	*ksɪa*	*sɔ*	aaaaa
31	drink	*diʔ*	*dɨnniet*	*diʔ*	*deʔ*	*teəŋ*	aaaab
32	dry	*rkʰiaŋ*	*riəŋkʰoŋ*	*raw*	*rhɪəŋ*	*raj, roh*	aabac
33	dull, blunt	*isiʔ*	*ɟosmoʲt*	*tʰla*	*len*	--	abcd*
34	dust	*pum-pum*	*pum-pum*	*dʒlpʰoʔ*	*tjəlpʰʊʔ*	*kərboh*	a*bbc
35	ear	*ʃkor*	*ləkur*	*ʧkɔr*	*tərəŋ*	*hjɔʔ*	aaabc
36	earth	*kʰɨndew*	*kmiəŋ*	*kʰndaw*	*pərθa*	*kətɛ*	abc*d
37	eat	*baːm*	*bɨnnəŋ*	*bam*	*bʊə*	*hap*	aaaac
38	egg	*pɨlleŋ*	*pɨlliŋ*	*plleiɲ*	*sʔi*	*kətəm*	aaabc
39	eye	*kʰmat*	*kʰmat*	*kʰmat*	*mat*	*ŋaj*	aaaab
40	to fall	*hap*	*eŋaj*	*hap*	*hərem*	*rar*	abacd
41	far	*ɟŋaj*	*ɟŋi*	*dʒŋai*	*ʧŋʊ*	*səŋaj*	aaaaa
42	fat-grease	*sŋaːʲd*	*ɨmmir*	*kʰlaŋ*	*ləʔɒt*	*kəmu*	abcde
43	father	*kpa*	*pa*	*pa*	*pa*	*kun*	aaaab
44	to fear	*ʃeptieŋ*	*tieŋ dait*	*teʲɲ*	*ktɪəŋ*	*jɔ*	aaaab
45	feather	*sner*	*snir*	*tʰawanɛr*	*θəbənɪar*	--	aaaa*
46	few	*kʰɨn-diat*	*tah-diat*	*kʰadʒʲak*	--	*bre*	aab*c
47	to fight	*jaʃoʔ / jadat*	*jamuʔ*	*jatʃɔʔ*	*jaʔ dat*	*taik*	aba*c
48	fire	*diŋ*	*ədoŋ*	*diɲ*	*ʃmen*	*ŋər*	aaabc
49	fish	*doʔkʰa*	*kʰa*	*dakʰa*	*hi*	*ka*	aaaaa
50	five	*san*	*san*	*san*	*ran*	*pʰən*	aaaaa
51	to float	*per*	*raŋ*	*pɛr*	*sper*	*plur*	abaac
52	to flow	*tuːʲd*	*sɨn-toːʲd*	*toːʲd*	*pɒr*	*hlaj*	aaabc
53	flower	*sɨntiew*	*sɨntew*	*sɲtu*	*khlʊə*	*poh*	aaabc
54	to fly	*her*	*kɨndej*	*pɲhɛr*	*pɪar*	*pər*	abacc
55	fog	*dum-lʔoʔ*	*nioŋ nia*	*lʔɔʔ kʰɲdaw*	*dʊm*	*aj*	abcad
56	foot	*kɟat*	*kɟat*	*kdʒat*	*nɪa*	*ɟɯŋ*	aaabc
57	four	*saw*	*saw*	*so*	*rɪa*	*pʰon*	aaaab
58	freeze	*ʃoʔ tʰaʔ*	*tiŋŋam / bɨnsier*	*tʰaʔ*	*tʰaʔ*	*krɔʔ*	abaac
59	fruit	*soʔ*	*suʔ*	*sɔʔ*	*sɒʔ*	*ple*	aaaab
60	give	*aj*	*ɨnnaj*	*e*	*ʔaː*	*dɛh*	aaaab
61	good	*bʰa*	*mɨrrʰiaŋ*	*bʰa*	*mɪət*	*laʔ*	*a*bc
62	grass	*pʰlaŋ*	*pʰlaŋ*	*pʰlaŋ*	*smʊt*	*kərban*	aaabc
63	green	*ɟɨrŋam*	*sɨŋiəŋ*	*jɪrŋam*	*ʧərŋam*	*ɲər*	aaaab
64	guts, intestines	*snier*	*snor*	*snɛr*	*nɒr*	*rɛŋ*	aaaab
65	hair	*sɲiuʔ*	*sɲiək*	*sɲʲoʔ*	*sʊʔ*	*huʔ*	aaaaa
66	hand	*kti*	*ktej*	*kti*	*taː*	*ti*	aaaaa
67	he	*u*	*ɟutuʔ / umi*	*u*	*u*	*ən*	aaaab
68	head	*kʰlieʔ*	*kʰliʔ*	*kʰleiʔ*	*kʰlɪa*	*kiŋ*	aaaab
69	hear	*sŋap*	*sŋu*	*sniaw*	*sãʔ*	*ɟu*	aaabc
70	heart	*kloŋ snam*	*kloŋ snam*	*klɔŋ snam*	*klɒŋ rnʊə*	*nuər*	aaaab
71	heavy	*heʔ*	*kenbaʔ / kʰɨnnia*	*kʰia*	*stʊʔ*	*ɟən*	abcde
72	here	*haŋ ne*	*haniʔ*	*heini*	*tineː*	--	aaaa*
73	hit	*tied*	*udaʔ*	*dat*	*dat*	*tum*	abccd
74	hold-take	*ʃɨm*	*tʰom*	*tʃɨm*	*lʊm*	*lɛ*	abacd

75	how	*kumno*	*naŋ net*	*kammɔn*	*kinjaʔ*	*kʰuj mɔ*	abacd
76	hunt	*beʔ (mraːd)*	*wuŋ na*	*lai siɛt dɔʔ*	*pətar*	*ɻɔm*	abcde
77	husband	*tŋa / lok*	*koraŋ*	*lɔk*	*lʊk*	*rəleh*	abbac
78	I	*ŋa*	*nə*	*ŋa*	*ŋə*	*ʔɔ*	aaaab
79	ice	*tʰaʔ*	*tʰaʔ-əlliʔ*	*tʰaʔ*	*tʰaʔ*	*je ge*	aaaa*
80	if	*lada*	*lede*	*lada*	*nimə*	--	aaab*
81	in	*ha*	*he*	*ha*	*ti*	*naə*	aaab*
82	kill	*pɨnjap*	*pɨnŋap*	*pɲjap*	*pənjip*	*piəm*	aaaab
83	know	*tip*	*heʔkən*	*tɪp*	*tʊʔ*	*nəp*	abaac
84	lake	*puŋ*	*puŋ*	*puŋ*	*sʊ*	*nɔŋ*	aaab*
85	laugh	*rkʰie*	*ɨllom*	*rkʰai*	*rʊ*	*jum*	abac*
86	leaf	*sla*	*sla*	*sla*	*sli*	*hla*	aaaaa
87	left side	*diaŋ*	*tɨmmiəŋ*	*tidiɛŋ*	*di par ta dɪəŋ*	*ʔi-ve*	aaaab
88	leg	*kɻat*	*kɻat*	*kdʒat*	*nɪa*	*ɻuŋ*	aaabc
89	to lie, deceive	*tʰok*	*tʰɨlloʲt*	*dʒlɛr*	*pənrʊʔ*	*cɔʔ*	abcde
90	live	*im*	*ɨnnim*	*ɪm*	*pʔem*	*im*	aaaaa
91	liver	*doʔnuːd*	*noːd*	*noːd*	*kθim*	*kərtɔm*	aaabb
92	long	*ɻroŋ*	*ɻirroŋ*	*dʒrɔŋ*	*kərʊŋ*	*hluŋ*	aaaab
93	louse	*ksi / ɻɨnreiŋ*	*sɨlliet*	*ksi*	*ksa*	*si*	abaaa
94	man-male	*ʃinraŋ*	*kʰonkoraŋ*	*cɳraŋ*	*tərma*	*ime*	aaabb
95	many	*bun*	*bon*	*bon*	*ʃibʊə*	*kun, bərcu*	aaaab
96	meat-flesh	*doʔ*	*meʲm*	*dɔʔ*	*dʊʔ*	*jəŋ*	abaac
97	mother	*kmie*	*gma*	*bei*	*maː*	*ma*	aaaaa
98	mountain	*lum*	*dom*	*lom*	*pdeŋ*	*sor*	aaabc
99	mouth	*ʃintur*	*gap*	*ktein*	*tkʊŋ*	*mur*	abcde
100	name	*kɨrteŋ*	*kɨrteŋ*	*prtuid*	*tvɪəŋ*	*ɻu*	a*bc*
101	narrow	*raʃiŋ / bakʰim*	*bakʰim*	*kʰɪm*	--	*ɔp*	a*a**
102	near	*ɻan*	*ɻŋan*	*dʒan*	*tjan*	*dət*	aaaab
103	neck	*rɨndaŋ*	*kraŋ*	*rdaŋ*	*rdaŋ*	*rəmɔŋ*	abaac
104	new	*tʰɨmmaj*	*tʰɨmmaj*	*tʰɱme*	*θmaː*	*kənmɛ*	aaaaa
105	night	*miet*	*sənŋu*	*meit*	*ləmaʔ*	*sum*	abaac
106	nose	*kʰmut*	*leumut*	*kʰmut*	*mərkɤ̃ŋ*	*muh, mur*	aaaba
107	not	*ɨm*	*ɨnɻi*	*ɱ*	*tʊʔ tə*	*kə*	aabbc
108	old	*rim*	*rim*	*rɪm*	*sərem*	*prim*	aaaaa
109	one	*uwej*	*uwew*	*wi*	*mi*	*u*	aaaab
110	other	*ki-wej*	*marber*	*kɔpsar*	--	*laj*	aa**b
111	person	*briew*	*brü*	*bru*	*tjəpreʊ*	*bi*	aaaab
112	to play	*leʔ kaj*	*kʰellaj*	*kɳdeiʔ ke*	*khirʊ*	*kəvɛʔ*	abacd
113	to pull	*tan*	*rɨnnieŋ*	*tan*	*pətɪaʔ*	*ruit, tut*	abacd
114	to push	*kʰɨnniaʔ*	*kɨncʰew*	*ŋiat*	*khən jit*	*con*	abacd
115	to rain	*slap*	*slap*	*slap*	*slaː*	*ɻuŋ*	aaaab
116	red	*saw*	*ənsaw*	*so*	*sɪa*	*kʰo, ni*	aaaab
117	right-correct	*dej*	*dew*	*tɔʔ*	*tʊʔ*	--	aabb*
118	right side	*mon*	*tɨm-mon*	*timun*	*di par ta mɒn*	*kʰwa*	aaaa*
119	river	*waʔ*	*por*	*waʔ*	*ʔam*	*om*	abacc
120	road	*linti*	*twar*	*sərɔk*	*sərʊk*	*deŋ*	abacd
121	root	*tɨnraj*	*tɨrraj*	*tʰeid*	*ʃɨt*	*riər*	aabba
122	rope	*tɨllaj*	*laŋnaj*	*tɭle*	*tərʊ*	*vər*	abacd
123	rotten	*pʔut*	*ɻillit*	*pʔut*	*khvi*	*əm*	abacd
124	to rub	*kɨrʃut*	*kɨrcʰut*	*kɽtʃut*	*kəntʃʊt*	*sut*	aaaaa
125	salt	*mluʔ*	*maluk*	*bloʔ*	*pnʊʔ*	*sɔʔ*	aaaab
126	sand	*ʃʔiap*	*cʰʔjap*	*tʃʔɛʌp*	*sərvʊ*	*saj*	aaabc
127	to say	*oŋ*	*ɨnnoŋ*	*ʔɔŋ*	*ʔɒŋ*	*dah*	aaaab
128	to scratch	*truːd*	*tɨrrut*	*tʃbɔt*	*kʰəbɒʔ*	*pɔʔ*	aabbc
129	sea	*duriaw*	*duriaw*	*duriaw*	*dʊrɪəʊ*	--	a*aa*
130	to see	*joʔi*	*muɻoʔ*	*pait*	*maʔ*	*ju*	aabca
131	seed	*sɨmbaj*	*ɻellej*	*sɱbe*	*tjʊsba*	*kəʔaŋ*	abaac
132	to sew	*suʔ*	*sɨnnek*	*sor*	*sʊ*	*ɻiŋ*	aaaab

K. S. NAGARAJA, P. SIDWELL, S. GREENHILL. 2013.
A Lexicostatistical Study of the Khasian Languages. *Mon-Khmer Studies*. 42:1-11

133	sharp	*nɐp*	*ɨntaʔ*	*nɛp*	*nɐp*	*ləm*	abaac
134	short	*liŋkot*	*tɨmban*	*tbien*	*ʧrit*	*ɛm*	abbcd
135	to sing	*rwaj*	*rɨŋwi*	*rwai*	*rvʊ*	*ŋir*	aaaab
136	to sit	*ʃoŋ*	*cʰoŋ*	*tʃɔŋ*	*ʃkɪə*	*məʔ*	aaabc
137	skin	*snieʔ doʔ*	*snieʔ mejm*	*sneiʔ*	*snɪəʔ*	*hur*	aaaab
138	sky	*bneŋ*	*brej*	*bneiŋ*	--	*pleŋ*	aba*a
139	to sleep	*tʰiaʔ*	*ɨŋŋiŋ*	*tʰiaʔ*	*θɪəʔ*	*ʔit*	abaac
140	small	*rit*	*doh-dit*	*kʰeiŋ*	*sbɪət*	*diət*	abcdb
141	to smell	*sma*	*ɨnnaw*	*sma*	*rʔɪəŋ*	*ʔur*	abacd
142	smoke	*tdem*	*ɨntʰak*	*tdɛm*	*tdem*	*tuk*	abaac
143	smooth	*ɟliʔ*	*ɟɨmpaj*	*jaliʔ*	*tjəlliʔ*	*kleət*	abaac
144	snake	*bseŋ*	*bseŋ*	*pseiŋ*	*psen*	*hiŋ*	aaaaa
145	snow	*jor*	*jor*	*tʰaʔ*	*ksɪəŋ məŋ*	--	a*bc*
146	some	*kʰɨndiat*	*taʔ-diat*	*kʰaɟiak*	*ʃitʃĩŋ*	*pərdi*	aabcd
147	to spit	*biaʔ*	*ɟɨrtʰew*	*mɲtʰu*	*pəθeʊ*	*bɛʔ*	abbba
148	to split	*pʰiaʔ*	*tʰɨillaʔ*	*pʰiaʔ*	*phit, khlɪəʔ*	*ploh*	abaac
149	to squeeze	*kʰem*	*kʰɨnnim*	*ksiʔ*	*ʧəpɪə*	*piət*	aabcc
150	to stab-pierce	*duŋ*	*daneŋ*	*duŋ*	*təndoŋ*	*bruh*	aaaab
151	to stand	*ieŋ*	*ñiəŋ*	*jeiŋ*	*rəŋ*	*ɟəŋ*	aaaaa
152	star	*kʰlur*	*kʰlor*	*kʰlor*	*khlʊə ʃmen*	*səmiŋ*	aaabb
153	stick (of wood)	*dieŋ*	*ədiəŋ*	*deiñ*	*pərnɪa*	*hviət*	aabcd
154	stone	*maw*	*maw*	*mo*	*ʃmɪa*	*mo*	aaaaa
155	straight	*biet*	*lɨmpʰar*	*beit*	*bɪt*	*pʰiəŋ*	abaac
156	to suck	*kɟit*	*kɨnɟok*	*bu*	*tjʊr*	*bu, but*	aabcb
157	sun	*sŋi*	*sŋej*	*sŋi*	*njʊŋa*	*səŋi*	aaaaa
158	to swell	*at*	*ɨnnat*	*ad*	*ʔat*	*gɯ*	aaaab
159	to swim	*ɟŋi*	*ɟɨnnaj*	*jɱpa*	*rɪəŋ*	*lɔj*	aabac
160	tail	*tdoŋ*	*kdoŋ*	*tdɔŋ*	*tdɒŋ*	*staʔ*	aaaab
161	that	*-ta / -tej*	*ga-tej*	*katai*	*ke/u tʊn*	*taj*	aaaba
162	there	*katʰie*	*gatʰoʔ*	*heitai*	*tʊ tʊn*	--	aabc*
163	they	*ki*	*gniʔ*	*ki*	*jə*	*ge*	abaca
164	thick	*rben*	*rɨmbin*	*rbɛn*	*rben*	*hət*	aaaab
165	thin	*staŋ*	*sɨntaŋ*	*staŋ*	*staŋ*	*hrer*	aaa*b
166	think	*pɨrkʰat*	*pɨrkʰat*	*pɽkʰat*	*pərkhat*	*tʰaŋ*	a***b
167	this	*-ne*	*ga-niʔ*	*kani*	*ke/u ne*	*ʔɯ*	aaaab
168	thou	*me / pʰa*	*mi / pʰe*	*me / pʰa*	*ɪəm (m) / ɪəhe (f)*	*mi*	aaaba
169	three	*laj*	*laj*	*le*	*la*	*ʔuəj*	aaaab
170	to throw	*kawaŋ*	*lɨntʰew*	*pakʰɔt*	*phədat*	*rup*	abcde
171	to tie	*teʔ*	*tɨnnak*	*kdɔʔ*	*kʊt*	*toʔ*	aabca
172	tongue	*tʰɨillieɟ*	*tʰɨlloʲt*	*tʰɽleiʔ*	*khlit*	*kərtaʔ*	aaaab
173	tooth	*bniat*	*moʲŋ*	*lʌmeiŋ*	*ləmen*	*hraŋ*	abbbc
174	tree	*dieŋ*	*diəŋ*	*deiŋ*	*tvɪa*	*he*	aaabc
175	to turn	*kɨilla*	*kɨilla*	*dɔŋ*	*khərvi*	*pən*	a*bcd
176	two	*aːr*	*ar*	*ar*	*ʔõə*	*ʔar*	aaaaa
177	to vomit	*prei*	*pɨrraw*	*prai*	*hərɒʔ*	*hur*	aaabc
178	to walk	*jaːʲd*	*dinniʔ*	*lai kdʒat*	*lɪa*	*pʰət*	abcd*
179	warm	*sʔaːʲd*	*ɨnʃɨt*	*cit*	*dɒt*	*kəʔur*	abbcd
180	to wash	*sait*	*sait*	*sait*	*ksi*	*kʰoj, kəta*	a*aab
181	water	*um*	*gum*	*um*	*ʔam*	*ʔom*	aaaaa
182	we	*ŋi*	*jew*	*i*	*ʔɪə*	*ʔɛ*	abccc
183	wet	*ɟʰieʔ*	*ɟimbaʲt*	*dʒheiʔ*	*tjəriəʔ*	*om pjo pjo*	abaac
184	what	*-ej*	*umet*	*ilɛʔ*	*i a*	*mɔ*	abcad
185	when	*lano*	*minnet*	*ɳnu*	*daŋ njaʔ*	*jam*	abac*
186	where	*haej*	*hanet*	*ʧeiwʌn*	*ti njaʔ*	*mɔ*	ab*cd
187	white	*lieʔ*	*əlliʔ*	*leiʔ*	*slaŋ*	*blɔʔ*	aaabc
188	who	*-no*	*jət*	*u/ka ji*	*u/ke ʔaː*	*paj*	abbcd
189	wide	*jar*	*ɨniər*	*jar*	*hɪaŋ*	*vah*	abcde
190	wife	*tŋa*	*kontʰaw*	*lɔk*	*kə lɒk*	*pənle*	abcde

K. S. NAGARAJA, P. SIDWELL, S. GREENHILL. 2013.
A Lexicostatistical Study of the Khasian Languages. *Mon-Khmer Studies*. 42:1-11

191	wind	*lʔer*	*lʔier*	*lʔɛr*	*srʊə*	*kur*	aaaab
192	wing	*tʰapŋiaŋ*	*tʰapnir*	*tʰawanɛr*	*θəbənɪar*	*peəŋ*	aaaab
193	to wipe	*ŋiad*	*innat*	*ɲiam*	*ʃʔɪəʔ*	*kʰut, kʰuit*	abcde
194	with	*bad*	*nam*	*wa*	*bəʔ*	--	abca*
195	women	*kɨntʰej*	*rawkmaw*	*kn̩tʰai*	*hənθa*	*ɪpən*	abaac
196	woods, forest	*kʰlaw*	*ləwtəp*	*kʰlo*	*kərmɪa*	*bri*	abacd
197	worm	*wieʔ*	*wiak*	*weiʔ*	*khvi*	*riər*	aaaab
198	ye > you (pl.)	*pʰi*	*pʰjaw*	*pʰi*	*hi*	*pɛ*	aaaaa
199	year	*snem*	*snim*	*snɛm*	*snem*	*sənəm*	aaaaa
200	yellow	*stem*	*sɨntim*	*stɛm*	*tŋʊə*	*teŋ*	aaabc